

## **Part V**

# **Nulling Interferometry**



## Chapter 16

# Nulling Interferometry and Planet Detection

EUGENE SERABYN

JET PROPULSION LABORATORY  
CALIFORNIA INSTITUTE OF TECHNOLOGY  
PASADENA, CALIFORNIA

### 16.1 Introduction

The small angular separation and high brightness contrast characterizing star-planet systems has thus far prevented the direct detection of planets beyond our solar system. If our solar system were viewed from a distance of 10 pc, a mere  $0.1''$  would separate the Earth from the Sun (Figure 16.1), while Saturn would lie  $1''$  out. Moreover, the Sun/Earth contrast ratio at visual wavelengths would exceed  $10^9$ . However, while the angular separations are fixed constraints, the contrast ratio is amenable to modification in at least two ways. First, the observational wavelength can be shifted into the infrared, where planetary thermal emission spectra peak (Bracewell, 1978; Angel *et al.*, 1986). Second, it is possible to alter the intrinsic contrast ratio with an optical approach capable of selectively dimming the star relative to its surroundings.

In the latter regard, two techniques are quite promising: coronagraphy and nulling interferometry. Because coronagraphy relies on physically blocking the incident starlight with a small obscuring spot located at a focal plane stellar image, useful coronagraphic imaging is possible only beyond several Airy radii from the on-axis stellar source. This approach is thus likely optimal primarily for the nearest stars. On the other hand, as is elaborated in the succeeding sections, nulling interferometry is expected to be effective only within the core of a telescope's point-spread function, and so can be employed for stars at greater distances, where a larger sample is available.

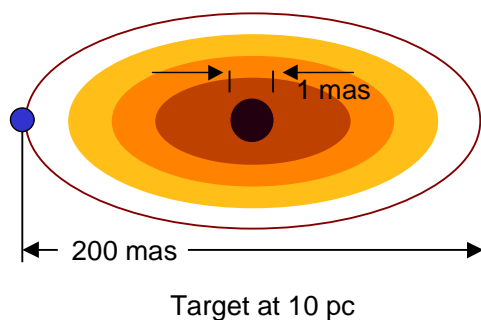


Figure 16.1: Schematic diagram of the inner part of our own solar system as seen from a distance of 10 pc. Included are the Sun, Earth, and zodiacal cloud. The centrally peaked zodiacal emission actually extends to radii well beyond 1 AU.

In this paper, an overview of the techniques of “nulling” interferometry is presented. As a concrete example, the operation and performance of fiber-coupled rotational shearing interferometers is described in some detail. However, it should be borne in mind that at this stage few of the suggested nulling approaches have been demonstrated experimentally, and so the optimal approach may not yet be identified.

## 16.2 Exozodiacal Light

Beside proximity and contrast, a third obstacle to direct exoplanet detection is the possible emission from “exozodiacal” dust grains congregating in and near the orbital plane of the target solar system (Figure 16.1). Such dust disks can be expected both by analogy with our own solar system’s zodiacal cloud (Reach *et al.*, 1995), and by extrapolation of the few known cases of very bright circumstellar disk emission around young main sequence stars (Backman and Paresce, 1993). In fact, due to the extended nature of such exozodiacal disk emission, a disk’s integrated thermal emission is likely to exceed even that of gas giant planets by orders of magnitude. For example, the integrated thermal emission from our own solar system’s zodiacal dust is about  $10^{-4}$  that of the sun’s luminosity at a wavelength of  $10\ \mu\text{m}$ , while the emission from Jupiter is two orders of magnitude lower, and the Earth’s emission yet another order of magnitude lower (Traub *et al.*, 1996; Angel, 1998; Beichman *et al.*, 1999; Serabyn *et al.*, 2000).

While observations of nearby stars have established the presence of dust in a few systems, the high contrast ratio involved means that the dust detected to date tends to be located at large distances from the central stars. It is thus largely colder dust at Kuiper-belt scales ( $> 30\ \text{AU}$  radius) which has been detected in these systems. Information on warmer dust at the smaller radial offsets commensurate with habitable zones is therefore almost nonexistent. Prior to conducting extensive searches for very weak emission from terrestrial planets around nearby stars, it would therefore be advantageous to establish the emission levels

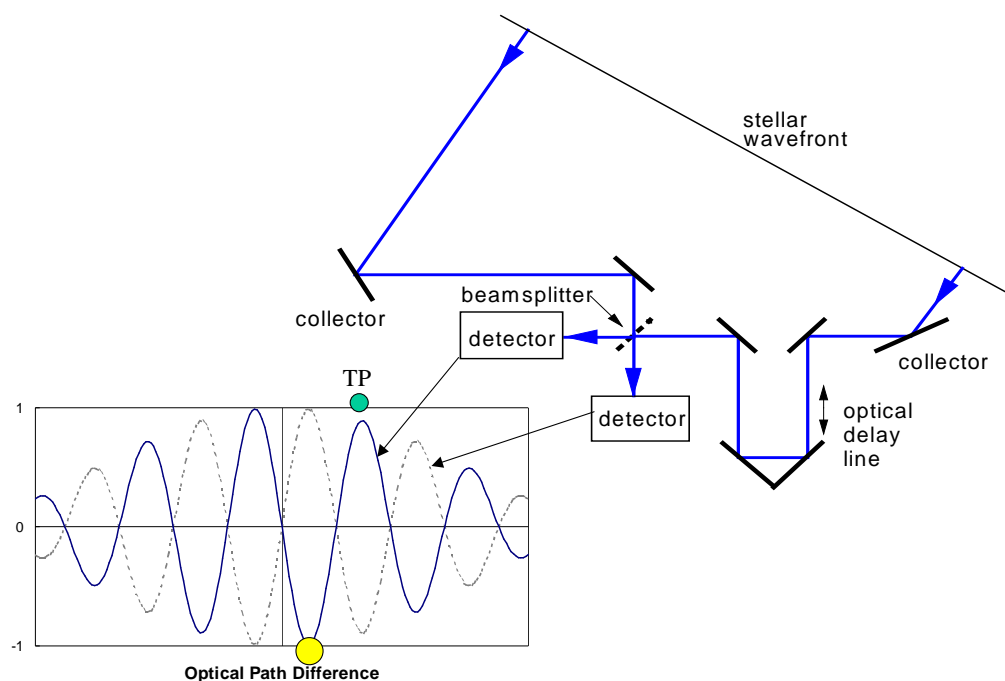


Figure 16.2: Schematic diagram of a standard stellar interferometer and the resultant fringe pattern at the two beam-splitter outputs. Conceptually the goal is to place the star at the bottom of a deep destructive fringe, and an accompanying terrestrial planet (TP) near the top of a constructive fringe.

from exozodiacal dust around these stars, and indeed, first-generation nulling experiments are being designed primarily to determine these exozodiacal emission levels. In particular, both the Keck Interferometer and Large Binocular Telescope nulling experiments are being designed to search for the warm dust at AU scales (Angel, 1998; Colavita *et al.*, 1998; Booth *et al.*, 1999; Serabyn *et al.*, 2000).

## 16.3 Nulling Interferometry Basics

The basic premise of nulling interferometry is conceptually quite simple: combine the light incident on a pair of telescopes so that at zero optical path difference (OPD) between the incident beams, the two electric-field vectors are exactly  $180^\circ$  out of phase, thus allowing near-perfect starlight subtraction. In the language of interferometry (Bracewell, 1978), a deep destructive fringe is to be placed across the star (Figure 16.2). However at the same time, off-axis emission from sources located near constructive fringe maxima can be transmitted through the system (Figure 16.2), and so even though the star is nulled to a deep level, appropriately situated planets are not attenuated greatly. The trick then is to devise an optical scheme which can null starlight of both polarizations and at all wavelengths across the passband of interest simultaneously.

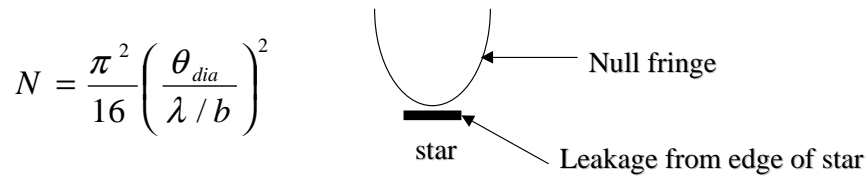


Figure 16.3: Cross-sectional view through the null fringe and star. The finite stellar diameter allows for light from the edges of the star to “leak” through the fringe pattern. The null depth,  $N$ , is the ratio of transmitted powers in the destructive and constructive states.

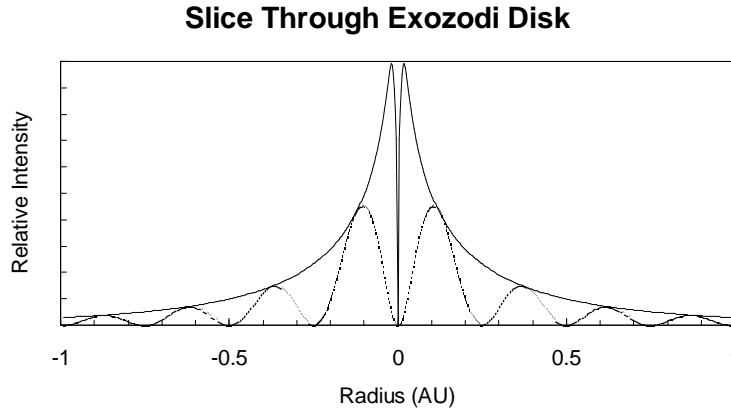


Figure 16.4: Cross-sectional view through an edge-on exozodiacal cloud at a distance of 10 pc, showing the emergent flux (solid curve), and the flux transmitted by the 10  $\mu$ m fringe pattern of the Keck Interferometer in the nulling mode (dotted curve).

The next section addresses the technical challenges which must be met in order for the potential of deep stellar nulling to be realized, but there is one fundamental null depth limitation which cannot be overcome, because it arises in a property of the source itself: the finite stellar diameter. For a uniform circular disk source, and a single pair of nulling telescopes, stellar radiation from points off the central axis of symmetry can leak around the central destructive fringe, as indicated in Figure 16.3. Thus, for a single baseline, the stellar cancellation cannot be very deep unless the baseline is short enough to spread the central destructive fringe across the entire stellar disk. For example, a sun-like G2 star at a distance of 10 pc has a diameter of 0.93 mas, implying that on the 85-m baseline between the two Keck telescopes, 10- $\mu$ m null depths (where the null depth is defined as the ratio of transmitted powers in the destructive and constructive states) for solar-equivalent stars are limited to about  $10^{-3}$ . At the same time, the off-axis fringes can transmit a substantial fraction of the exozodiacal and exoplanetary radiation. Figure 16.4 shows the effect of the Keck-Keck fringe pattern (of period 24 mas) on an exozodiacal cloud like that of our own solar system’s, at a distance of 10 pc: approximately half the arriving exozodiacal flux is transmitted by the nulling interferometer’s off-axis fringes.

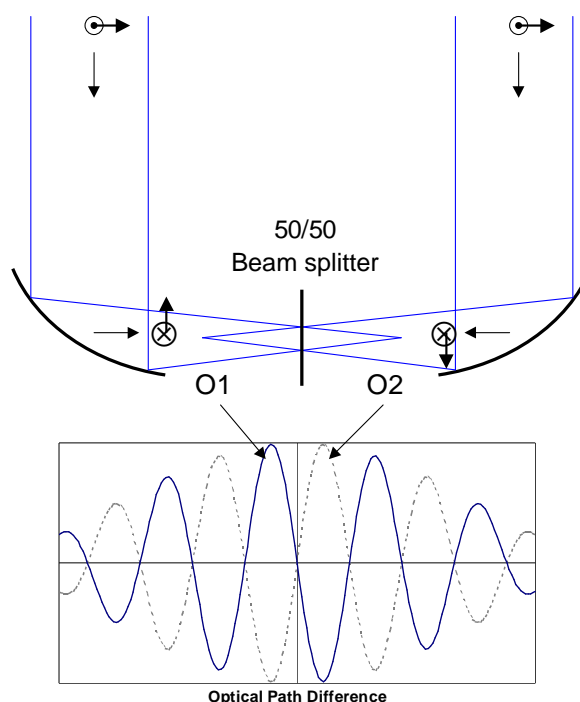


Figure 16.5: Original nulling scheme proposed by Bracewell & MacPhie (1979).

For the nulling approach to be effective, the stellar cancellation must be very deep, on the order of a part in  $10^6$  for planet detection, and a part in  $10^{3-4}$  for exozodi detection (Beichman *et al.*, 1999). Of course the null depths given refer to the net nulls integrated across the passband. Such a deep achromatic null fringe cannot possibly be provided by a standard stellar interferometer, because as Figures 16.2 and 16.5 show, the inherent symmetry of such systems implies that at zero OPD, the two beam-splitter outputs are equivalent, with each receiving half the incident power. To effect cancellation, a finite OPD must therefore be introduced. An offset of a quarter of the average wavelength in the passband will optimize the cancellation (Figures 16.2 and 16.5), but since a quarter wavelength is clearly a chromatic quantity, it is evident that all wavelengths cannot cancel simultaneously at a non-zero OPD setting. Indeed, this is the reason the original nulling scheme (Figure 16.5) of Bracewell and MacPhie (1979) fails to provide a very deep null. (Of course, there is an even more serious problem with this scheme, in that the fringes in the two polarization states actually cancel each other completely for such a perfectly symmetric system).

A more achromatic approach is therefore needed, the ideal case being completely wavelength-independent cancellation at zero OPD. What is thus desired is a set of fringes which are complementary to the case of a standard laboratory Michelson interferometer (Figure 16.6). Such an ideal achromatic destructive fringe at zero OPD (solid curve in Figure 16.6) can be generated by subtracting, instead of adding, the two incident electric fields. In fact, by con-

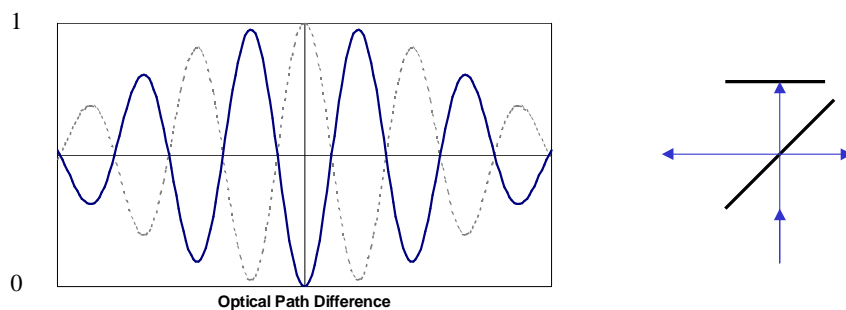


Figure 16.6: Standard laboratory Michelson interferometer, and the associated fringe pattern at the output (dotted line). For deep achromatic nulling, what is needed is an inverted fringe pattern (solid line), with achromatic cancellation zero OPD.

servation of energy such subtraction applies at a Michelson interferometer's complementary output. However, as this output sends the light back to the input port, it is not readily accessible. To enable the necessary subtraction at a more accessible output port, two approaches have been considered: a relative flip of the electric-field vectors (Diner, 1990; Shao, 1990), which is intrinsically achromatic, and a phase retardation approach (chromatic by design) in which light at each wavelength is delayed by exactly the distance necessary for all waves to arrive exactly  $180^\circ$  out of phase (Woolf and Angel, 1998).

Two approaches have been proposed for implementing a field-flip, both based on variants of rotational shearing interferometers. In the first approach (Diner, 1990; Shao, 1990), the flat end mirrors in the two arms of a Michelson interferometer are replaced by a pair of rooftop mirrors, oriented so that they appear orthogonal when viewed in projection in the common output beam (Figure 16.7). This rotational shearing interferometer concept will be described more fully below, but in brief, each rooftop flips one component of the incident field (Figure 16.8), so that a relative flip of the total electric-field vector,  $\mathbf{E}$ , is induced by the pair of rooftops. This approach clearly has the advantage of relying solely on flat mirrors.

In a second configuration (Baudoz *et al.*, 1998a,b), the basic Michelson interferometer is again modified by replacing the two flat end mirrors, but this time with focusing cat's eye assemblies (Figure 16.9). In one of the interferometer's two arms, the cat's eye secondary is a flat located at the focus of the primary mirror, while in the second arm, the flat secondary is replaced by a curved mirror which reflects the light back on itself before focus is reached. Thus, a focus is present in only one of the interferometer's two arms. Since passage through focus introduces an achromatic phase shift of  $\pi$  radians into a beam of light (Born and Wolf, 1980), upon recombination at the second beam splitter pass the two beams emerge from the interferometer with their electric-field vectors flipped relative to each other. Because of the different secondary mirrors in the two cat's eyes, the optical system is not quite symmetric, but for focal ratios greater than about 10, the limitations incurred thereby are insignificant.



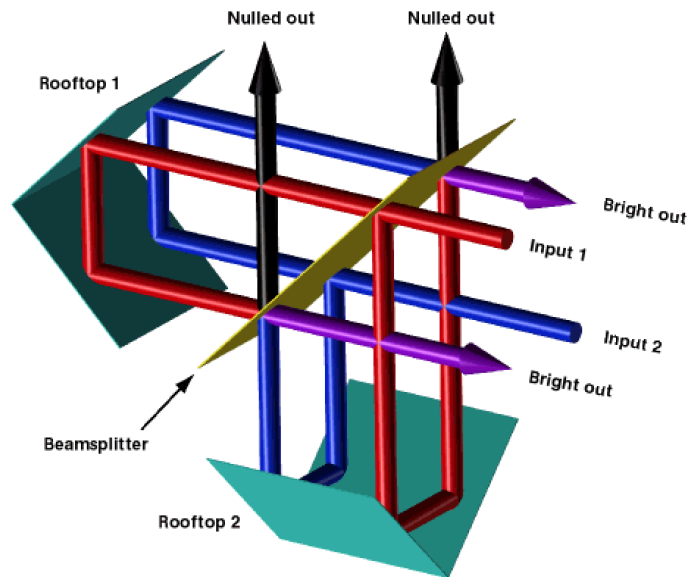


Figure 16.7: Orthogonal-rooftop-based nulling interferometer. Each rooftop flips only that field component which is perpendicular to the rooftop apex line, resulting in a net field flip between the two arms. In this configuration, the two inputs lead to two nulled and two “bright” outputs.

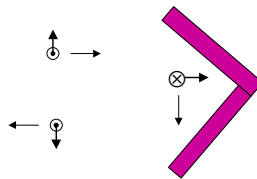


Figure 16.8: Side view of a rooftop (dihedral) mirror, illustrating the fact that only the  $E$ -field component normal to the rooftop apex emerges flipped in direction.

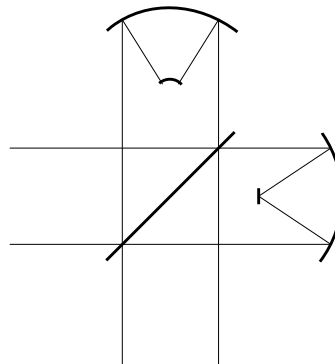


Figure 16.9: Cat's eye-based nulling interferometer. A focus is present on the flat cat's eye secondary in the right arm, but the upper arm has no focus. By virtue of passage through focus, an extra phase shift of  $\pi$  radians is present in the right arm.

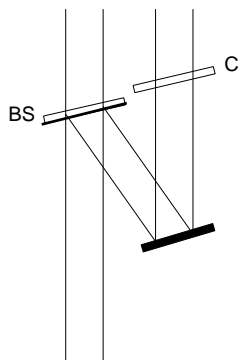


Figure 16.10: Phase retardation scheme. The dielectrics comprising the compensator are designed to add a quarter wavelength of phase across the band. The ideal, lossless beam splitter introduces a second  $\pi/4$  of phase.

Finally, instead of simply flipping the  $\mathbf{E}$ -field vectors, there is the alternative of introducing an equivalent wavelength-dependent phase retardation (Hinz *et al.*, 1999; Morgan and Burge, 1999). In this approach, the dielectric indices of a beam splitter/compensator pair are carefully selected to delay each wavelength in the passband by the requisite half wavelength. Since in this approach the beam splitter is used in single-pass, the goal is to introduce an achromatic  $\pi/4$  phase shift at the beam splitter, with another  $\pi/4$  introduced by the compensator (Figure 16.10).

The phase retardation approach is thus quite different from the field-flip approach, and one essential difference in particular is worth noting: in both of the field-flip approaches the beam splitter is used in double pass inside a small beam-combining interferometer, while in the phase retardation approach, the beam splitter is used in single pass (Figure 16.11). Thus, in the first case, the two input beams are both modulated by similar beam-splitter reflection-transmission products, and so they retain their original intensity ratio, while in the latter case, one beam is modulated by the beam splitter's transmission coefficient and the other by its reflection coefficient. In the first case, power balance between the two beams is automatically maintained (for equal input powers), independent of the detailed beam-splitter properties, thus allowing for high-accuracy subtraction. In the latter case, power balance requires equality of the beam-splitter reflection and transmission coefficients.

## 16.4 Symmetry and Stability Requirements

The on-axis cancellation of stellar radiation requires a very symmetric and stable optical system. In particular, for a given polarization state, the two beam intensities, electric-field rotation angles (Figure 16.12), and phase delays must all be matched to  $2\sqrt{N}$  (Serabyn, 2000), where  $N$  is the null depth as defined earlier. Thus for a net null depth of  $10^{-6}$ , the allowable intensity mismatch (the deviation of either beam intensity from the mean), the relative rotation angle, and the phase difference must all be individually less than about

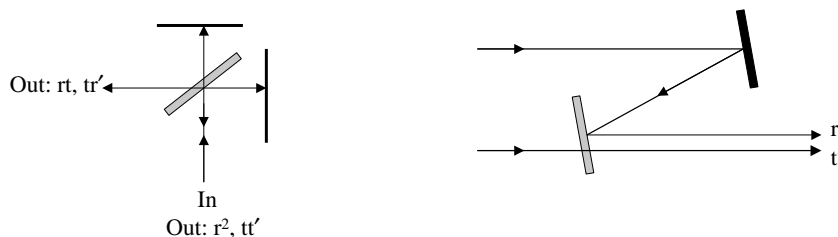


Figure 16.11: Comparison of beam splitter roles in the two nulling approaches. In the field-flip approach (left) the beam splitter is used in double pass, and the powers are automatically balanced; in the phase-retardation approach (right), one input field is multiplied by the beam-splitter reflection coefficient, and the other input by the transmission coefficient.

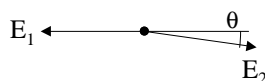


Figure 16.12: In the plane normal to the propagation direction, the angle between the two E-field vectors must be within  $2\sqrt{N}$  radians of  $\pi$ .

$10^{-3}$  (the latter two in radians). This fine degree of balancing can only be achieved if a suitable means of adjusting the relevant parameters exists.

The rotation angle can be adjusted fairly accurately by adjusting the accompanying image rotation. As this is a static quantity, it is the easiest of the three to deal with. The relative beam intensities can be modified by introducing slight relative pointing offsets, or by introducing adjustable small-area obscurations across the centers of the beams. Possibilities in the latter category include a small scissor-like device (Figure 16.13, left), or a small Venetian-blind-like rotating vane (Figure 16.13, right). Finally, the relative phases (path delays) can be adjusted by means of precision optical delay lines, as are typically used in stellar interferometers, but with accuracies on the order of 1 nm to allow  $10^{-6}$  nulling at  $\lambda = 10 \mu\text{m}$ . Of course, also needed for the latter is a control algorithm, i.e., a means of sensing non-zero phase errors. The simplest control approach, and the only one which has been experimentally demonstrated to date, is cavity-length dithering (Serabyn *et al.*, 1999a), but this approach loses efficacy as the photon flux is diminished. Possible alternatives include control of one nuller output by means of the second (Serabyn, 1999), control by means of light in a non-nulled waveband (Hinz *et al.*, 1999), and control of one polarization component by means of the second. All three of these approaches require experimental demonstration. A final possibility is of course laser metrology.

Moreover, these conditions must be met simultaneously for both polarization components, at every point across the aperture, and for all wavelengths in the band. Thus it is also critical to avoid effects such as “differential birefringence”: differing s-p phase delays between the two interferometer arms. The problem is illustrated graphically in Figure 16.14, where it can

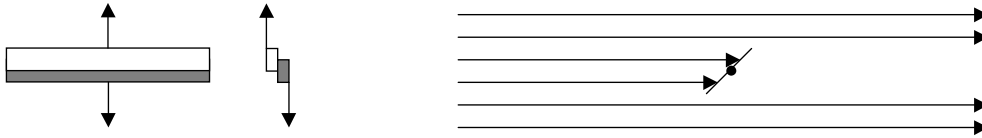


Figure 16.13: Two possible intensity modulation schemes, both based upon a linear obstruction of variable width crossing the center of the aperture of the beam. On the left is a scissor-like mechanism, and on the right is a rotating vane which resembles a Venetian-blind.

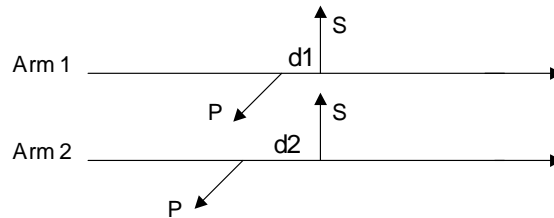


Figure 16.14: Diagram illustrating “differential birefringence”,  $d_2 - d_1$ . In this example, the s-waves have been phased up, but the p-waves are out of phase by  $d_2 - d_1$ . Alternatively, by adding a delay of this magnitude to arm 1, the p-waves could be phased up, but at the cost of the s-waves going out of phase by the same amount.

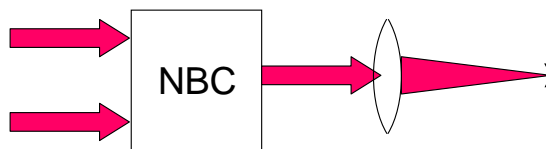


Figure 16.15: Schematic showing the use of a single-mode spatial filter to clean up the wavefront and deepen the null.

be seen that in the presence of differing s-p phase delays, either one of the two polarization states can be phased up, but not both simultaneously. The limit on the allowable differential birefringence is  $4\sqrt{N}$  (Serabyn, 2000).

Satisfaction of the condition of simultaneous nulling across the full beam aperture brings a further experimental requirement, as this condition effectively means translating the aforementioned phase error requirement into a surface accuracy requirement. As a phase accuracy of  $10^{-3}$  radians at a wavelength of  $10\ \mu\text{m}$  implies a surface accuracy of order 1 nm, it is clear that this goal cannot be met even with the finest optical surfaces available. Indeed, if  $S$  is the beam Strehl ratio, attainable null depths would be limited to  $\sim 1 - S$ , or  $10^{-2}$ . Another approach is therefore necessary. However, this limitation can be overcome (Shao, 1991; Shao and Colavita, 1992; Ollivier and Mariotti, 1997) by means of spatial filtering in the output focal plane (Figure 16.15), as the center of the focal-plane point spread function is the Fourier transform of the *average* aperture-plane field. Thus the focused nuller output beam needs to be passed through a single-mode spatial filter, i.e., a single mode fiber at optical wavelengths, or a pinhole in the mid-infrared. This effectively limits operation to a single diffraction-limited mode of the telescope aperture, and so an analogy to single-dish, single-detector radio astronomy is not inappropriate.

Finally, the need to satisfy all of the above conditions across the passband of interest calls for a minimization of dispersion. In particular (Serabyn, 2000), the mean square phase dispersion across the passband also cannot exceed  $2\sqrt{N}$ . This implies mainly careful matching of the beam splitter/compensator pair, although in laboratory experiments, the injection of the light must also be carried out with minimal dispersion. The effects of atmospheric dispersion must also be considered in this regard.

## 16.5 Rooftop-Based Rotational Shearing Interferometers

The use of a fiber-coupled rotational shearing interferometer (RSI) as a nulling beam combiner bears further examination. As pointed out in Figure 16.8, a single rooftop mirror acts to reverse only one component of the incident field—that component perpendicular to the rooftop apex. With two orthogonal rooftop mirrors located behind a beam splitter, as in Figure 16.7, both polarization components are then flipped relative to each other (Figure 16.16). In addition to flipping the  $\mathbf{E}$ -fields, the rooftop mirrors also shear the return beams symmetrically across the rooftop centerlines (Figure 16.17), so that two input beams yield two nulled and two bright (or constructive) output beams (Figure 16.7), all of which are separated from the input beams. Thus, in comparison to the phase-retardation approach of Figure 16.10, the number of output beams is doubled. The RSI approach thus appears somewhat more complicated, but on the other hand, it cleanly separates the field-flip and phase-delay issues, and also brings with it the ability to implement a pathlength control scheme in which one nulling output can be used to control the second (Serabyn, 1999).

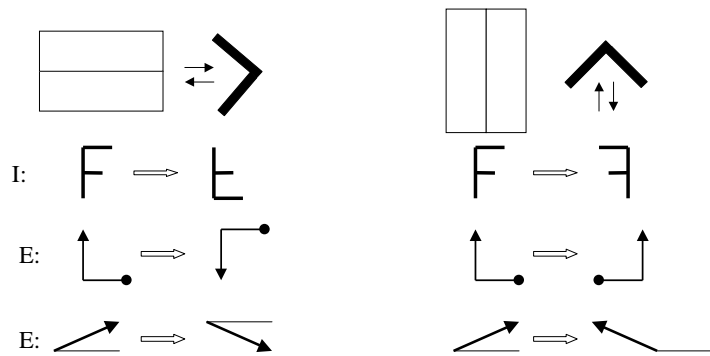


Figure 16.16: Illustration of the action of two orthogonal rooftop mirrors. I refers to the image, and E to the E-field.

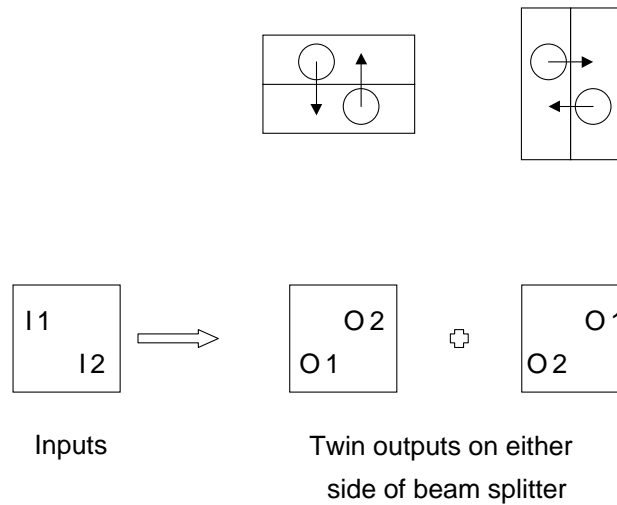


Figure 16.17: Illustration of the lateral beam shear obtained in a beam combiner based on an orthogonal-rooftop RSI.

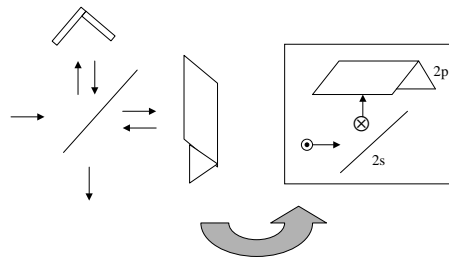


Figure 16.18: Diagram illustrating the fold mirrors needed to symmetrize the two arms of an orthogonal rooftop-based RSI.

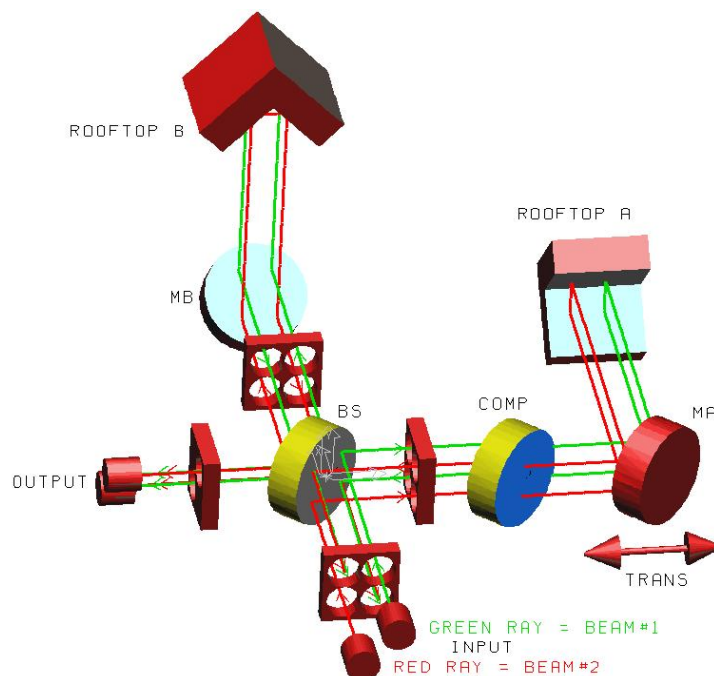


Figure 16.19: Optical layout of a polarization-compensated rooftop-based RSI. The two  $45^\circ$  fold mirrors, MA & MB, are needed to obtain complete symmetry between s- and p-reflections. The aperture plates with the four holes are guides for the eye.

However, one problem with the basic orthogonal-rooftop RSI shown in Figure 16.7 is that a significant asymmetry exists: light of a given polarization state (s- or p-plane polarized at the beam splitter) undergoes two s-plane reflections at one of the rooftop mirrors, but two p-plane reflections at the other. Thus incident light in the two polarization states would emerge with nonzero and opposite s-p phase delays. Luckily, it is possible to symmetrize the situation by inserting properly oriented  $45^\circ$  fold mirrors prior to the rooftops, as in Figure 16.18, so that light in each arm of the modified RSI sees two s-plane and two p-plane reflections, all at the same  $45^\circ$  angle of incidence (Shao, 1991; Diner *et al.*, 1991). This results in the overall optical layout shown in Figure 16.19.

## 16.6 Experimental Results

A fiber-coupled polarization-compensated rotational-shearing interferometer of the type described in the last section has been built at JPL (Figure 16.20) in order to test the potential of the approach. To date, the experiments at visible wavelengths have proven quite successful, verifying essentially all aspects of the fiber-coupled RSI approach except for dual-polarization operation. Indeed, after many cycles of component and environmental improvements, null depths of a part in  $10^5$  are now achieved regularly with a narrowband (0.5%) visible-wavelength laser diode source (Figure 16.21), and white light nulls of order a part in  $10^4$  have been achieved for 10% bandwidth, single-polarization thermal (red) light